PARTIAL LEAST SQUARES IS TO LISREL AS PRINCIPAL COMPONENTS ANALYSIS IS TO COMMON FACTOR ANALYSIS.

Wynne W. Chin
University of Calgary, CANADA

ABSTRACT

The decision of whether to use PLS instead of a covariance based structural equation modeling technique such as LISREL for causal modeling can be assisted by looking at the differences between principal components analysis and common factor analysis. Through such a process, this paper outlines the need for PLS users to shift from merely estimating model parameters to that of including measures of predictive relevance. Unless the communality is high and the indicators per construct are large, the PLS parameter estimates for construct loadings will likely have a homogenization and overestimation bias. Conversely, the structural paths tend to be underestimated.

Commentator's Biography

Wynne W. Chin is Associate Professor in Management Information Systems at the University of Calgary. He received an AB in biophysics from U. C. Berkeley, an MS in biomedical engineering from Northwestern University, and an MBA and Ph.D. in computers and information systems from The University of Michigan. He is currently completing the development of PLS-Graph - a Windows based software for performing Partial Least Squares analysis. His research interest involves the impact of individual attitudes and social networks on user acceptance of new information technology. His other research interests include methodological issues in information systems research, group support systems, and end user computing.

Mailing Address: Faculty of Management, University of Calgary, Calgary, Alberta T2N 1N4, Canada; Phone: (403) 220-3732; Fax (403) 282-0095; E-mail: chin@acs.ucalgary.ca on Internet.

This is a copy of my original submission that was published as Chin, W. W. (1995). Partial Least Squares is to LISREL as principal components analysis is to common factor analysis. *Technology Studies*, 2, pp. 315-319. The final publication was less complete than this document since it omitted the abstract, made several mathematical typos, and left out two references. Otherwise, everything else remains the same.

I am pleased to see Barclay, Higgins, and Thompson (hereafter BHT) produce this article which provides a much needed introduction of the Partial Least Squares (PLS) approach for technology researchers. While the paper suggests that researchers interested in learning about PLS should complete a thorough review of principal components analysis, path analysis, and OLS regression, I would extend this by strongly suggesting the need to understand the differences between common factor analysis and principal component analysis (Widaman, 1993; Velicer and Jackson, 1990). The arguments concerning the pros and cons for choosing one over the other can by analogy be used to compare PLS with covariance based techniques such as LISREL. In both situations, issues to be considered include the epistemic relationship between data and theory (which includes the issue of factor indeterminacy), factors that affect population parameter estimates (i.e., sample size, communality, number of indicators per factor, and sample data distribution), and the objective for the analysis (prediction versus model confirmation).

In general, the primary difference between factor analysis and components analysis (thus equally true for LISREL and PLS) is whether one wishes to explicitly model unique factor variances for manifest variables. The common factor analysis model may be expressed in matrix notation as:

$$C - U^2 \cong F_U F_U^T = F_R \Phi F_R^T = C^*$$

where C is the matrix of correlations among manifest variables, U^2 is a diagonal matrix of estimates of unique variance, F_U and F_R are the unrotated and rotated factor matrix respectively, Φ is the matrix of factor intercorrelations, and C^* is the reproduced correlations among the manifest variables.

In the case of components analysis, rather than representing the shared variance among a set of manifest variables, the total variance is represented in a reduced-dimensional form. Thus, the standard matrix equation is represented as:

$$C \cong F_U F_U^T = F_R \Phi F_R^T = C^*$$

As U^2 approaches zero, the factor loadings and factor correlations using both methods approach identity. Yet, as BHT note, valid variance of an typical indicator tend to be 50 to 83%. The remaining 17 to 50% unique variance represent a combination of valid specific variance (i.e., reliable variance of the indicator that is not shared with the latent variables in the analysis) and measurement error. As the proportion of unique variance increase due to specific variance, the greater the difference will be for results obtained from a components based method such as PLS contrasted to a factor analytic LISREL approach. This distinction has been shown analytically for models with orthogonal factors (e.g., McArdle, 1990) and empirically for models with oblique factors (Widaman, 1993).

McArdle (1990) takes this distinction to form several statements contrasting components analysis and factor analysis. By analogy, I suggest these statements should also be considered for PLS and LISREL.

1) Choice of method depends on whether the researcher wishes to maximize the multivariate variance of manifest variables or in reproducing the population parameters that underlie all the covariances?

The goal of PLS is primarily to estimate the variance of endogenous constructs and in turn their respective manifest variables (if reflective). Models which I have developed yielding significant jackknife statistics can still be invalid in a predictive sense. Thus, I would suggest the focus should be shifted from only assessing the significance of parameter estimates (i.e., loadings and structural paths) to that of predictive validity. Predictive sample reuse technique as developed by Geisser (1974) and Stone (1975) represent a synthesis of cross-validation and function fitting with the perspective "that prediction of observables or potential observables is of much greater relevance than the estimation of what are often artificial constructs-parameters" (Geisser, 1975, p. 320).

The PLSX sub module of Lohmöller's PLS 1.8 program provides this functionality as part of a blindfolding algorithm (1981, p. 5.9-5.12). In blindfolding, portions of the data for a particular construct block (i.e., indicators by cases for a specific construct) are omitted and cross-validated using the estimates obtained from the remaining data points. This procedure is repeated with a different set of data points as dictated by the blindfold omission number until all sets have been processed. A resulting relevance measure is thus obtained for the endogenous construct in question. This relevance measure is generally more informative than the R^2 and the average variance extracted since the latter two have the inherent bias of being assessed on the same data that were used to estimate its parameters. Alternative sample reuse methods employing bootstrapping or jackknifing have yet to be implemented.

2) LISREL is superior to PLS on mathematical grounds.

This point refers to the fact that LISREL is a population based model for estimating loadings and structural path estimates. As BHT note, only under the joint condition of large sample size and large number of indicators per factor will the estimate of the factor loadings and structural path estimates approximate that of the LISREL estimate. Otherwise, the loadings in a PLS analysis tend to be overestimated and the structural paths, conversely, underestimated (Dijkstra, 1983, p. 86). An examination of the component versus common factor distinctions will also suggest that communality represent yet another factor. Widaman's (1993) simulation study, for example, show that 3 to 7 indicators are needed for accurate assessment of component loadings for orthogonal factors under the condition of high communality in the population (i.e., .80 loadings for all indicators). If the underlying population loadings are less saturated at .60, 10 to 18 indicators are needed to obtain an accurate assessment. Otherwise, the estimated loadings tend to deviate from the true values by more than 10 percent (e.g., actual .60 loading is estimated as .757, actual .80 loading is estimated as .872, and actual .40 loading is estimated as .663). Common factor analysis, on the other hand, is able to extract the true underlying population parameters.

With the tendency for overestimation of loadings, the corresponding structural path estimates are conversely underestimated. Widaman's (1993) simulation show that under oblique factor conditions where the population correlation among factors is set at .50, component analysis result in an estimate of .421 when an equal loading pattern for the population is set at .80. The estimate drops to .182 when actual loadings are .40. Common factor analysis, on the other hand, tend to slightly underrepresent the loadings (i.e., .77 instead of .80 and .35 instead of .40) with a corresponding over representation of the correlation at .537 under high communality and .645 for low communality.

Equally important is the tendency of a components analysis to homogenize the loadings for a construct when the actual pattern is varied. If a three indicator factor model have population loadings of .8, .6., and .4, Widaman's study show that a component based analysis using oblique factor rotation result in estimations of .771, .769, and .709 respectively. In the orthogonal condition, the estimates are .825, .782, and .642 respectively.

Finally, item loadings under components analysis can vary depending on the test battery of items. Widaman (1993) performed two components analysis on a test indicator with a specified population loading of .6. In the first analysis, two other items were set with population loadings of .80. In the second analysis, the loadings were .40. Beyond the expected overestimation of the population loading of .60, the estimate differed between the two conditions resulting in .808 and .702 respectively. Thus, item loadings estimates from a components analysis is not invariant under different test batteries for the same common factor. Common factor analyses using iterated communalities and orthoblique rotation, on the other hand, were able to reproduce the population values.

Thus, superiority of LISREL over PLS refers to the ability to estimate the underlying population parameters. As noted in statement one, this becomes less of a concern if the objective is to account for multivariate variance in a predictive sense. Further, under conditions of low theoretical knowledge which BHT suggest is true for much of technological based studies, the more conservative estimate of a model's structural paths may be more appropriate. PLS estimates for misspecified models where non-significant structural paths are suggested will, by default, not be as large as the equivalent LISREL estimates.

3) LISREL is superior to PLS on statistical grounds.

This statement is relatively contentious and depends on the perspective of the researcher. The reverse statement suggesting that PLS has better statistical sampling properties than LISREL can equally be made. BHT take this second position by noting that PLS makes no distributional assumption regarding the data. The statement that PLS makes no distributional assumption relate to the asymptotic efficiency of the OLS estimator. Yet, due to the nature of the PLS algorithm, the construct score estimates are biased and are only consistent under the condition of high communality, appropriate number of indicators per construct, and increasing sample size. Nonetheless, because PLS is a limited information estimation procedure, an appropriate sample size tends to be much smaller than that needed for a full information procedure such as LISREL.

BHT also note that PLS produce component scores while LISREL is inherently indeterminate. Yet, as noted in statement 1 and elaborated in statement 2, the value of having PLS component scores need to be articulated. Under the goal of parameter estimation, it is not clear whether PLS weights and loadings (and thus PLS scores) are as generalizable across different samples as those obtained from LISREL. Even with distributional violation, the Maximum Likelihood estimation procedure for LISREL can be quite robust and may possibly, as alluded to in statement 2, produce better estimates of the population parameters. The generalizability issue for multiple group comparisons need also be considered. LISREL provides a statistical basis using a chi-square test for multiple group comparison. The generalizability of PLS scores for group comparisons have yet to be determined.

Predictive relevance, on the other hand, is a different issue that should be further explored. For example, the use of a mean loss function on the holdout data in a sample reuse procedure can be a viable means for choice selection of indicators. This would be in line with the data analytic exploratory depiction of PLS given by BHT. Likewise, the selection of various structural models can equally be assessed in this manner.

4) PLS is superior LISREL on practical grounds.

PLS is computationally more efficient than LISREL in the same sense as a components analysis is faster than a Maximum Likelihood factor analysis. BHT clearly articulate this point by noting that large models consisting of many indicators and factors can be estimated in a matter of minutes. In contrast, LISREL estimation time increases dramatically as the number of indicators increase.

In summary, an understanding of the issues related to choice of component versus common factor analysis can provide a basis for choosing PLS or LISREL as an analysis technique. BHT clearly articulated that the aim of LISREL is to estimate causal model parameters whereas PLS is to maximize variance explained. To understand this point further, one need only look at the component/factor analytic distinctions.

In both cases, the choice of the indicators and theoretical model still represent a necessary condition. When using PLS, low theoretical knowledge does not necessarily imply a researcher's inability to define constructs nor the nomological network in which these constructs reside. Instead, it likely depicts an exploratory stage where a researcher is testing an ad hoc model with newly developed items. If new scale items are created (likely with an average communality of .60), it would seem prudent to have a much larger set of indicators per construct (at the level of approximately 12 to 16) in order to obtain an accurate estimate of the structural paths. For PLS, there is less ability to disentangle poor indicators within a test battery for a particular construct. Given the homogenization and overestimation bias, I would be cautious of accepting items with loadings less than .80. Finally even if one's model consists of constructs with high levels of internal consistency, in order to be consistent with the causal-predictive goal of PLS, greater focus should be paid on the predictive relevance of a model.

References Cited

Dijkstra, T. (1983). Some comments on maximum likelihood and partial least squares methods. Journal of Econometrics, 22, 67-90.

Geisser, S. (1974). A predictive approach to the random effect model. Biometrika, 61 (1), 101-107.

Geisser, S. (1975). The predictive sample reuse method with applications. Journal of the American Statistical Association, 70, 320-328.

McArdle, J. (1990). Principles versus principals of structural factor analyses. Multivariate Behavioral Research, 25 (1), 81-87.

Lohmöller, J. B. (1981). LVPLS 1.6 program manual: Latent variables path analysis with partial least-squares estimation. Munich: University of the Federal Armed Forces.

Stone, M. (1975). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society, Series B, 36 (2), 111-133.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. Multivariate Behavioral Research, 25 (1), 1-28.

Widaman, K. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? Multivariate Behavioral Research, 28 (3), 263-311.